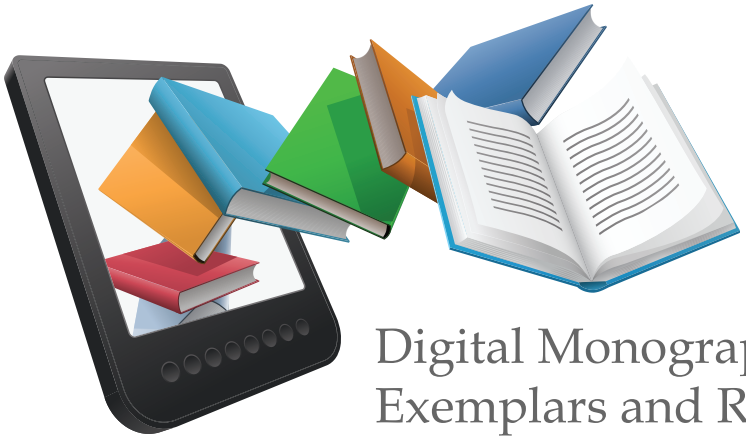
Technical Landscape Exemplars and Recommendations



JISCpub

Digital Monographs: Technical Landscape Exemplars and Recommendations

Technical Landscape Exemplars and Recommendations

EDiNA

JISC

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

The Digital Monograph Technical Landscape study (#jiscPUB) was a six month thinktank set up by the JISC in the first half of 2011 to explore the potential value that the use of the ePUB specification could bring to the Higher and Further Education sector if further adopted in UK Universities.

The thinktank members comprised Theo Andrew, Liza Daly, Peter Sefton, Emma Tonkin, Harsh Khatri and David F. Flanders. The final report was authored by Liza Daly.

The Digital Monograph Technical Landscape study (#jiscPUB) was supported by JISC as part of it's Repository Infrastructure Programme.

The quality of this report was greatly improved by a number of reviewers including John Norman, Mark Toole and Paul Walk.

EXECUTIVE SUMMARY

EXECUTIVE SUMMARY

Though digital publishing of journals and academic content has been flourishing for decades, the purchase and consumption of electronic books has only recently reached mass adoption in North America and the United Kingdom. Extrapolating from current trends, digital books are set to overtake print books in the global trade book marketplace; unit sales of ebooks already exceed those of print in many segments in the United States. Disintermediation of the publishing supply chain is occurring at all levels: from publisher, to agent, to authors.

To remain relevant to an audience that is increasingly consuming media in digital format only, scholarly authors and publishers must facilitate digital-first publishing techniques that are open, aid in discovery, and are highly accessible and device-independent.

This report aims to describe some historical perspective on electronic publishing, leading up to why the "ebook revolution" has happened in the 2010s when it had failed to take hold before. We will describe some of the details of how digital books are authored, both in a scholarly context and in general ebook production terms. We have included in-depth coverage of the unexpected outcomes of ebook distribution, including issues of rights, royalties, copyright, academic impact, and the implications of limiting access to and reproduction of digital books. Finally, we review findings from a number of ebook pilot programs conducted in U.S. universities, and draw from the work done by the rest of the "JISCpub" team in uncovering possible future work that could be actionable and relevant to a scholarly publishing audience, with a goal towards providing better tooling for both authors and readers of scholarly works.

RECOMMENDATIONS IN BRIEF

RECOMMENDATIONS IN BRIEF

Based on the work of the JISCpub thinktank these past six months we put forward the following recommendations to you for consideration. These recommendations are explained in full including suggested work that might be taken forward to advance these recommendations. The report itself provides the overall landscape and context for why we feel these recommendations are valid. Accordingly, the following recommendations we put forward are:

1. Rich full-text semantic search tools for scholarly ebook collections.
2. Tools for generating or traversing highly-specific stable citations.
3. Development of a pilot to produce student theses with high-engagement linked-data content.
4. Plugins or add-ons to provide simple, ebook output for popular word processing tools.
5. Improved workflows for authoring attractive, accessible, standards-based mathematical notation in ebooks.
6. Development of an ereading system with an emphasis on scholarly annotation and research-gathering.
7. Provisions to train and share scholars interested in digital publishing
8. Aggregate ebook services for authors and university presses.
9. Maximize use of orphan works
10. Community resources for institutions with digital collections

INTRODUCTION

INTRODUCTION

Sales of books have nearly exceeded overall sales of paperback books in the United States.

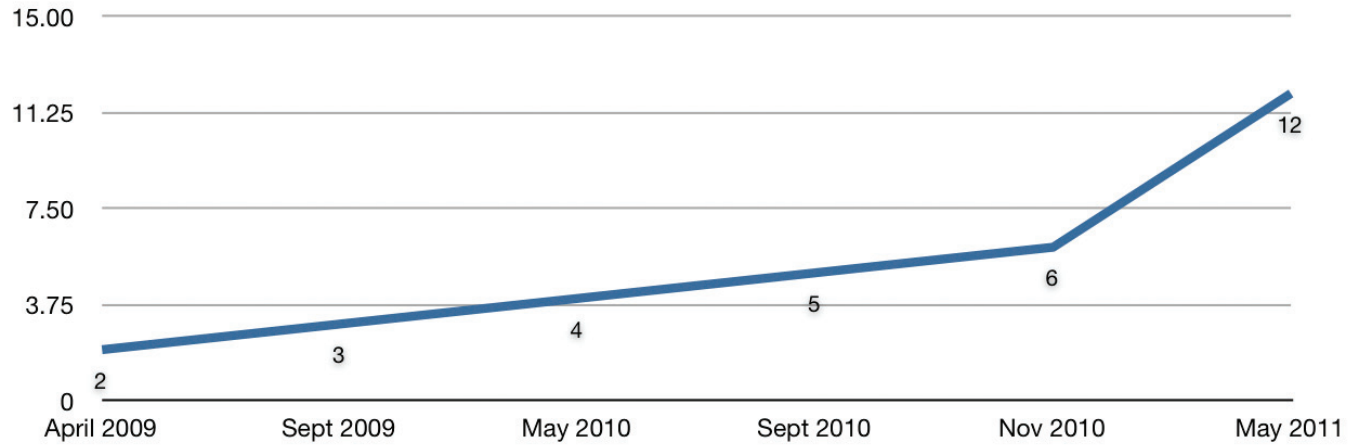
TABLE 1. TRADE EBOOKS SALES IN THE UNITED STATES

	April 2010	April 2011	Percentage change
Hardcover		$58.11M	
Paperback		$67.56M	
Ebooks	**$17.16M**	$44.15M	**+157.5%**

Source: The Association of American Publishers, June 2011
[http://www.publishers.org/press/38/]

While all major ebook retailers offer the ability to read their content on smartphones, tablets and computers, much of this sales growth is being fed by a sharp increase in ownership of dedicated ereading devices:

Percentage of US household with dedicated ereading devices



Source: Pew Internet Project, June 2011
[http://pewinternet.org/~/media//Files/Reports/2011/PIP_eReader_Tablet.pdf]

What caused this tremendous growth when earlier efforts at establishing an ebook ecosystem had failed?

Why digital publishing now?

Strictly speaking, there have been viable electronic publishing markets, especially in academia, since the late 1980s. Highly structured and regularized document formats such as those used in the legal and medical professions demonstrated early that there was tremendous value in providing large datasets of frequently-updated material in a digital form. Typically reference database products share a number of common factors that sped their adoption as electronic manifestations:

1. They have constrained formatting and markup requirements, meaning that content is not highly variable from instance to instance. All legal briefs or medical case studies are fundamentally similar to each other and can be easily modeled.
2. There exists a strong interest in searching across a collection to uncover material unknown to the searcher. For example, Lexis is a mature legal research product whose entrance into the market place long precedes the web, but is still a critical tool for the legal community today.
3. Reference database products have traditionally been bundled with the computing infrastructure, including simple computer terminals or more robust workstations, and reliable network connectivity. Compare the typical home computing environments in the 1980s and early 1990s to those provided by universities, and it is obvious why academic (journal) digital publishing matured early.
4. Reference products have greater value represented by the collection in aggregate than any number of individual discrete documents in the collection.

Early attempts at establishing a digital publishing ecosystem for trade (popular) book channels ran up against a number of challenges in the time period between the introduction of the early Palm Pilot ereading software (circa 1997) and until the release of the first Kindle (November, 2007):

1. High variability and uncertainty in ebook formats and encodings [http://en.wikipedia.org/wiki/Comparison_of_e-book_formats#Supporting_Hardware] Of the formats available, many popular formats such as PDF were ill-suited for use on mobile devices.
2. A limited set of material meant that early digital book stores provided far less value than an average print book store. Shoppers would be frequently frustrated in a search for titles and eventually abandon digital book shopping.

3. Portable reading devices required tethering to computers to make purchases, limiting usefulness "in the field."
4. Fiction content, particularly, does not derive obvious commerical value from being searchable or presented in aggregate (though shows tremendously potential for humanities researchers).

Early ereading devices (including desktop computers) provided little in the way of improvements on the reading experience. Screens on personal digital assistants (PDAs) like the Palm Pilot were either low-contrast monochrome or low-resolution color. Font choices were limited and font rendering was jagged and unsightly, requiring large font sizes for legibility and reducing the effective viewport size. Color screens could provide color imagery, but again low resolution inhibited the usefulness of embedded graphics, and efficient methods of encoding digital video had not been invented.

Battery life was a severe constraint on utility, a problem largely solved (somewhat paradoxically) not by increasing the efficiency of portable batteries but by radically shrinking the size and weight of all other components. An iPad derives its justly-praised 10 hours of usage time by being effectively a giant battery fronted by a thin pane of glass. Electronic ink devices solve the problem via their extremely low-power displays and use of plastic rather than glass.

Perhaps most importantly, early digital book ecosystems did not offer sufficient improvements in accessibility. Single-purpose devices were too expensive for wide audiences. The use of PDF as an ebook format inhibits the usefulness of screen readers, and cannot be reflowed into smaller, portable devices. Formats had not yet matured to provide sophisticated text/audio syncing as is provided by DAISY Talking Books [http://www.daisy.org/daisy-technology] or the EPUB 3 Media Overlay format [http://idpf.org/epub/30/spec/epub30-mediaoverlays.html].

The revolutionary aspects of even the first Kindle relied on providing solutions to the two most critical adoption deterrents: size and quality of catalog, and wireless delivery. Amazon had key advantages in bootstrapping the digital book ecosystem over earlier players. It has strong ties to a technophilic, book-buying audience — the first Kindle sold out within 6 hours [http://www.engadget.com/2007/11/21/kindle-sells-out-in-two-days/]. Amazon was able to leverage its considerable computing power and experience in "the cloud", and negotiate favorable contracts from device makers and, critically, wireless carriers who

were required to provide 3G always-on networking. Amazon solved the content problem by drawing from its strong (if somewhat contentious) relationships with publishers and providing its own print-to-digital conversion services when publishers did not yet have experience in creating ebooks themselves.

Once the electronic book industry was seeded, several factors have contributed to its continuing acceleration:

1. As print retail locations flounder and increasingly close altogether, buyers are driven to the web to make purchases. Since most major online retailers sell both print and digital books, even print buyers become progressively more exposed to the idea of reading digitally.
2. As more software vendors enter the marketplace, the cost of digitising material has fallen. These vendors quickly gain "on the ground" experience in producing ebooks, and quality has risen over time. As of 2011, the overall attention to detail provided to digital books still lags far behind print due to this reliance on external suppliers and inexperience within publishing houses, but as relationships between suppliers and publishers mature, editorial and technical quality rises.
3. As the number of digital book retail channels consolidates, the proliferation of formats and form factors begins to condense. It is now understood that EPUB serves as the universal "reflowable" format, [See Table 8] even though a different format is used internally on the Kindle.

Consistent with the pattern shown by the early reference databases, the titles which have demonstrated the greatest digital adoption have been those preferred by readers who want a large set of choices. Genre fiction, consumed quickly and in volume, frequently sell more units in digital form [http://www.straight.com/article-402353/vancouver/genrefiction-fans-quell-hunger-ebooks]. This is especially true for romance novels [http://www.rwa.org/cs/the_romance_genre/romance_literature_statistics/industry_statistics], in which the relative anonymity of digital books provides an unexpected value. In the United States, best-sellers are now routinely being sold primarily as ebooks rather than hardcover [http://www.nytimes.com/2010/07/20/technology/20kindle.html]. Though price is a major factor here, the relative disposability of ebooks is also an enticement when acquiring the book of the moment.

Digital publishing in academics

The move towards digital publishing in an academic context has taken place largely at the level of distribution. Aggregating journal articles and some book content (typically monographs) has been well-

established. Some university presses now produce academic texts with broad appeal for sale in ebook form. University of Michigan [http://press.umich.edu/ebooks/] sells PDF-format ebooks directly, some under a "rental" program of time-limited access, and Kindle-format books on Amazon.com [http://press.umich.edu/ebooks/kindle/index.jsp] with direct links from their own site. An entire series on digital culture [http://www.digitalculture.org/] is available under a Creative Commons license and readable for free online with tools from their MPublishing division [http://www.lib.umich.edu/mpublishing].

Indeed, some university presses adopted distribution via the Kindle early [http://www.insidehighered.com/news/2008/06/24/kindle]. Oxford University [http://www.amazon.co.uk/s/ref=nb_sb_noss?url=search-alias%3Ddigital-text&field-keywords=Oxford+University+Press%2C+UK] continues to publish book-length content in political science, cognitive science, and history briskly. BiblioVault [http://www.bibliovault.org/about.epl], a service of the University of Chicago Press, provides digitisation services and digital distribution to more than 80 scholarly presses. A service like BiblioVault, which helps to aggregate content across multiple very small presses, enables digital file production for scholarly presses that might otherwise not have the internal capability or resources.

Pressure to increase production of digital material is increasing due to some additional factors unique to educational publishing:

1. Escalating print textbook costs drive piracy of poorly-scanned digital copies of copywritten material. This indicates an untapped market for less expensive, more highly-targeted educational texts that are difficult to pirate and also provide pedigodical value beyond plain text and images.
2. Comfort with electronic search and retrieval means that students increasingly search *only* for material available in full and online. Works which are available in print only may be increasingly overlooked. Even works which are available via federated search systems in libraries are at a disadvantage relative to content found on the open web; proprietary databases require additional levels of authentication and often unique searching skillsets.
3. A need for academic publishers to counter the proliferation of dubious sources on the internet with authoritative works that are nevertheless congruent with emerging student needs and usage patterns.
4. Emergence of open access as a preferred or even mandated distribution method.
5. New ebook formats and channels like EPUB 3 [http://idpf.org/epub/30] or Inkling

[http://www.inkling.com/] provide options for pedogogical materials that would be impossible in printed form.

Though many of these factors are at present more pressing for textbook publishers than distributors of scholarly material, it is important to remember that the next generation of scholars will be educated in a context of increasingly-digital learning materials. Their expectations for ease of discovery, format-shifting, mobile access, and multimedia exemplars will extend beyond e-textbooks used at the undergraduate level.

HISTORICAL PERSPECTIVES

HISTORICAL PERSPECTIVES

Structured markup and academic authoring

Markup languages may be *presentational* or *semantic* (or a mix of the two, but generally one markup languages specializes in either even if it provides both capabilities). These two dimensions of a marked up text exist within a larger context of the viewport. Historically this viewport was fixed; first as a physical page, then as a virtual page: a PDF that may be printed and therefore should follow the locally-agreed upon page dimensions (A4 vs US letter).

Though the term "page" is still widely used even in digital publishing with full awareness that it is not a fixed size, it may be better to use the term "viewport." Related to but not entirely identical to the physical display size of a screen, the viewport is the visible area presented to the reader.

Why structured markup?

At first glance, TeX, XML and other structured markup would appear to be highly user-unfriendly. It does not resemble authoring tools such as Microsoft Word, which employ a "what you see is what you get" (WYSIWYG) modality of presenting a digital representation of the final text.

Instead, structured markup more closely resembles computer code. It tells a publication system exactly what the text *means*, rather than how it looks. While this is not appealing to most authors, the precision afforded by this approach is appealing to many researchers, especially in science and technology. Even some humanities scholars can be attracted to this hyper-specific approach, as seen in projects such as the massively complex Text Encoding Initiative [http://www.tei-c.org/index.xml].

Presentational languages such as TeX use a processing model in which the viewport size is known at document creation time. So while there may appear to be superficial similiarity between TeX-like markup languages and SGML-derived forms, the implications for layout choices available to the author are quite different. The layout "intelligence" in a traditional print process is with a human: the typesetter or book designer. In TeX, the layout process happens iteratively between the rendering system and the author: the rendering system interprets the markup; the author validates it, then revises the markup accordingly.

In SGML-derived markup languages such as XML (and by extension, all modern digital book formats), the author is no longer present at rendering time. The rendering engine must make layout choices based on hints available in the markup and contextual information from the device: its viewport size and dimensions, bit depth and quality of the display, even the time of day or physical location of the device.

DIGITAL BOOK LANDSCAPE TODAY

DIGITAL BOOK LANDSCAPE TODAY

Marketplace successes

Introduction of Kindle

In early 2005, Amazon quietly acquired the French ebook company Mobipocket [http://www.seattlepi.com/business/article/Retail-Notebook-Amazon-buys-e-book-software-1171055.php], in response to the slow dismantling of software and services provided by early ebook technology vendors, including Microsoft. By 2006, Amazon ceased selling all ebook formats other than Mobipocket, and independent, public release of new versions of the Mobipocket format and readers ceased.

The first hardware Kindle was officially released on November 19, 2007. It was an instant success, measured not only relative to the lacklustre sales of ebook products that preceded it, but in its own right.

While Sony had introduced a line of moderately popular eink devices before the Kindle was released, they lacked a companion ebook store with significant depth of titles, and the Sony devices had no networking capability at all; the device had to be connected to a physical computer to buy or transfer ebooks.

Formalization of EPUB/OEB

Many of the early ebook formats, including Mobipocket and Microsoft's LIT format, were based on the Open e-Book Publication Structure (OEB) [http://www.xml.com/lpt/a/646], a lightweight standard that described how an ebook could be packaged. OEB stopped short of defining all aspects of ebook formatting. The initial proliferation of incompatible OEB-derived formats caused its sponsoring organization, the International Digital Publishing Forum, to charter an effort to define a fully-specified ebook standard with a goal of becoming a universal format with rich accessibility. This resulted in the EPUB standard, released in September 2007, two months before the Kindle 1.

This first release of EPUB had several design goals:

1. Provide a standard for ebook markup that was unencumbered by patents or other legal restrictions on open use.
2. Be clearly specified by a diverse committee of ebook stakeholders, including publishers, device-makers, and accessibility proponents.
3. Re-use existing technologies wherever possible, including XML, XHTML, and accessibility standards.
4. Be implementable in a range of devices including low-memory hardware devices.

Sony adopted EPUB support quickly and discontinued its proprietary ebook format [http://www.nytimes.com/2009/08/13/technology/internet/13reader.html]. The Barnes & Noble Nook followed suit, as did Kobo and a number of other ebook retailers (some of whom did not survive the initial flourishing of ebook technologies).

EPUB is sometimes described as "a website in a box;" its technology stack is heavily borrowed from web technologies. This allowed a number of ebook readers to be developed quickly using web browsers as base platforms (examples include entirely web-based systems like Ibis Reader [http://ibisreader.com/] and Monocle) [http://monocle.inventivelabs.com.au/]. Nearly all ereading software found on tablets and phones uses a web-based component. EPUB's familiar technology platform has been perhaps the largest factor contributing to its swift adoption.

Strong growth and industry backing (plus its basis in familiar tools) meant that when technology companies entered the ebook marketplace, they chose EPUB as well. Steve Jobs mentioned the format by name when launching the iPad on-stage; Google acquired one of the oldest EPUB technology companies in 2010 and launched their own ebook platform using the format.

Ebooks and Digital Rights Management

Initial commercial adoption of EPUB was driven largely by the availability of a digital rights management (DRM) add-on provided by Adobe Systems. Adobe Content Server saw its earliest use in Sony devices, and later was licensed for use in the Kobo Reader and Barnes & Noble nook. To date, Apple is the only major retailer that sells EPUB books directly to US/UK consumers using a DRM mechanism not licensed by Adobe.

EPUB itself does not specify or endorse any particular digital rights management scheme, a feature for which it has been criticized. Though it would be theoretically possible to define an "open" DRM mechanism that could provide measure of content protection that was not owned by a particular vendor, there is only limited precedent for that in other digital media. DVD encryption, for example, is specified and regulated by an industry standards body; though it is worth noting that this encryption scheme has been broken and compliance with it is no longer legally enforced [http://w2.eff.org/IP/Video/DVDCCA_case/20040122_eff_pr.php].

Redistributing copywritten content is of course illegal in all countries with copyright legislation, including the UK [http://www.copyrightservice.co.uk/copyright/p01_uk_copyright_law]. Fair use guidelines in the UK allow for limited copying for "research purposes", and even in the United States (which does not recognize academic usage), copies of copywritten content can be made to accommodate use by individuals with disabilities [http://www.loc.gov/nls/reference/factsheets/copyright.html]. In both countries such exemptions are often managed by non-profits such as Bookshare [http://www.bookshare.org/?c=en_GB].

Producing tools to circumvent electronic access control mechanisms is illegal in the United States under the terms of the Digital Millenium Copyright Act (1998) [http://thomas.loc.gov/cgi-bin/bdquery/z?d105:H.R.2281:] in the United States. The related European Union InfoSoc Directive [http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:32001L0029:EN:HTML] goes further, prohibiting individuals from using such tools. Thus anti-DRM tools are illegal to produce in most Western countries, and in the EU and elsewhere using such tools is also illegal.

Despite the legal prohibitions, Adobe's ebook DRM (ADEPT) was broken in 2009. Software to remove copy protection from ebooks using ADEPT or the variant employed by Barnes & Noble are now widely available on the internet. The Kindle format, which is not EPUB, uses its own form of ebook encryption; that too has been broken and programs to remove the encryption are trivial to discover.

Citing piracy concerns, in November 2011 Penguin removed access to their catalog through the Overdrive/Amazon partnership with public libraries, prompting outrage from librarians [http://www.forbes.com/sites/alexknapp/2011/11/22/librarians-protest-penguin-decision-to-limit-e-book-lending/]. As this was a wholesale reduction in the content available through a previously-negotiated contract, the reaction is understandable.

It is also unclear whether the response was truly related to ebook privacy, or larger considerations about ebook distribution contracts. Other than isolated incidents like the above, the book industry has not seen widespread targeting of downloads of pirated ebook content as has occurred in the music and film industry.

DRM and library usage

Overdrive began providing ebook lending services to libraries in 2002, but limited content and format proliferation prevented wide adoption. Overdrive adopted EPUB in 2008; by 2011, the most widely-reported problem with ebook lending through Overdrive services is tremendous waiting lists.

Digital books do not suffer physical wear and tear and eventual decline (though changes in technology and format boom/bust cycles are often far shorter than the lifespan of a physical object). In 2011 Harper Collins announced an intention to limit circulation of individual ebook titles to 26 checkouts [http://www.libraryjournal.com/lj/home/889452-264/harpercollins_caps_loans_on_ebook.html.csp].

Librarian response to this was vehement [http://librarianbyday.net/2011/02/25/publishing-industry-forces-overdrive-and-other-library-ebook-vendors-to-take-a-giant-step-back/]. While some authors may perceive a benefit from reduced circulation (assuming that readers may instead purchase copies), others, especially authors who publish in series or deal in academic topics, depend on visibility in library collections to lead to future sales or professional recognition through citations.

Book lending on a global scale

The Internet Archive has used the same underlying technology platform as Overdrive for ebook lending, under their Internet Archive Lending Library [http://www.archive.org/details/lendinglibrary] brand. As of this writing, approximately 10,000 items were available for lending through the Internet Archive lending library, with a one-user, one-book lending model.

The Archive's lending policy is controversial; the organization views itself as a universal library, not restricted to geographic regions or communities of users. It has been careful [http://go-to-hellman.blogspot.com/2010/07/internet-archive-sets-fair-use-bait.htm] to select titles which are of minimal commercial viability, including those that are out of print. Indeed, the Archive's catalog of books are largely those that fall into the legal gray area of "orphan works" [http://en.wikipedia.org/wiki/Orphan_works]—those books which are legally in-copyright in a given region but for whom the copyright holder is not readily identifiable.

The problem of orphan works and digital availability of out-of-print books is especially acute in the academic community, where publications are potentially of interest to scholars indefinitely: long after print

US sales trends show continued growth in 2011

Sales growth for specialized devices continues on an upward trend.

TABLE 2. EREADER SALES 2010-2011

	Fourth Quarter, 2010	First Quarter, 2011
eReaders Worldwide:	6.0 million	3.3 million
2011 forecast:	**2.9 million**	**16.2 million**

Source: Sales data from IDC, 2011 [http://www.idc.com/getdoc.jsp?containerId=prUS22933011]

Relative to the initial flowering of many different tablets and eink devices in 2009-2010, 2011 has seen a consolidation around the "big players", with Barnes & Noble showing showing strong growth against the still-dominant Kindle (IDC, 2011) [http://www.idc.com/getdoc.jsp?containerId=prUS22933011]. It is unclear at this time whether this is due to increasing consumer preferences for a multipurpose color device like the Nook Color, or because of Barnes & Noble's continued powerful brick and mortar presence in the US.

TABLE 3. EREADING DEVICES ARE CONSISTENTLY FALLING IN PRICE

	Release date	US Price at release (in GDP)	US Price in July 2011 (in GDP)
Kindle	Nov 2007	$242	$69
Nook	Nov 2009	$157	$72
Kobo	May 2010	$90	$78

Prices are cited for the least-expensive model available directly from the retailer.

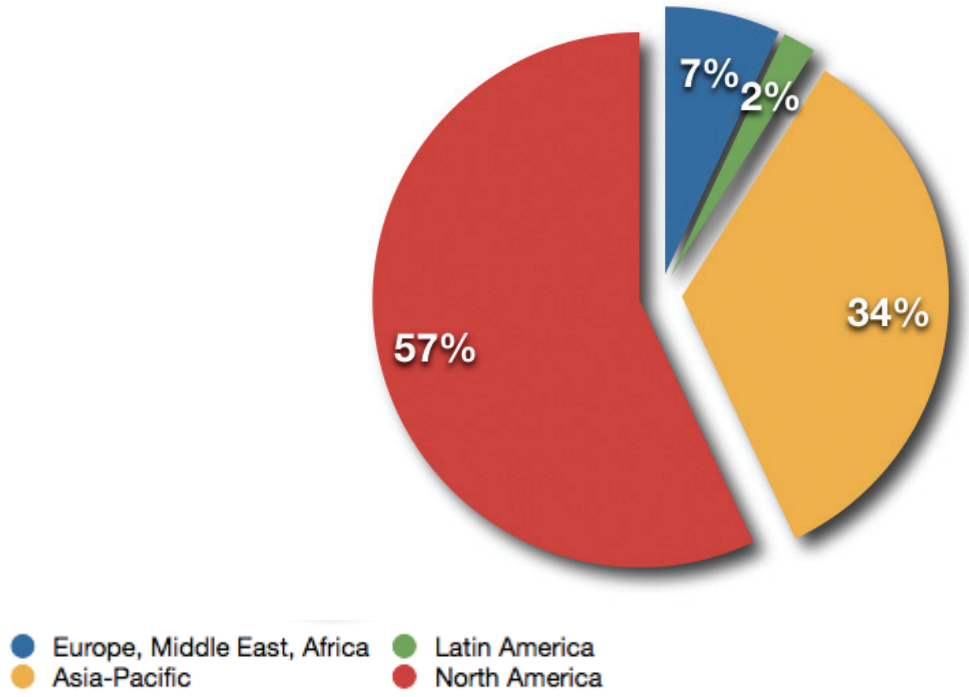
Trends in Europe and Asia

To date, no bookseller outside of the United States has made the kind of large-scale investment in ereaders as Barnes & Noble or Kindle. Indeed, early partnerships such as between Waterstone's and the Sony Reader [http://www.thebookseller.com/news/waterstones-reveals-sony-reader-sales.html] provoked some strong initial sales, but steep price wars ensued when Amazon opened its UK Kindle store [http://www.thebookseller.com/news/w-h-smith-slashes-prices-e-book-war-intensifies.html].

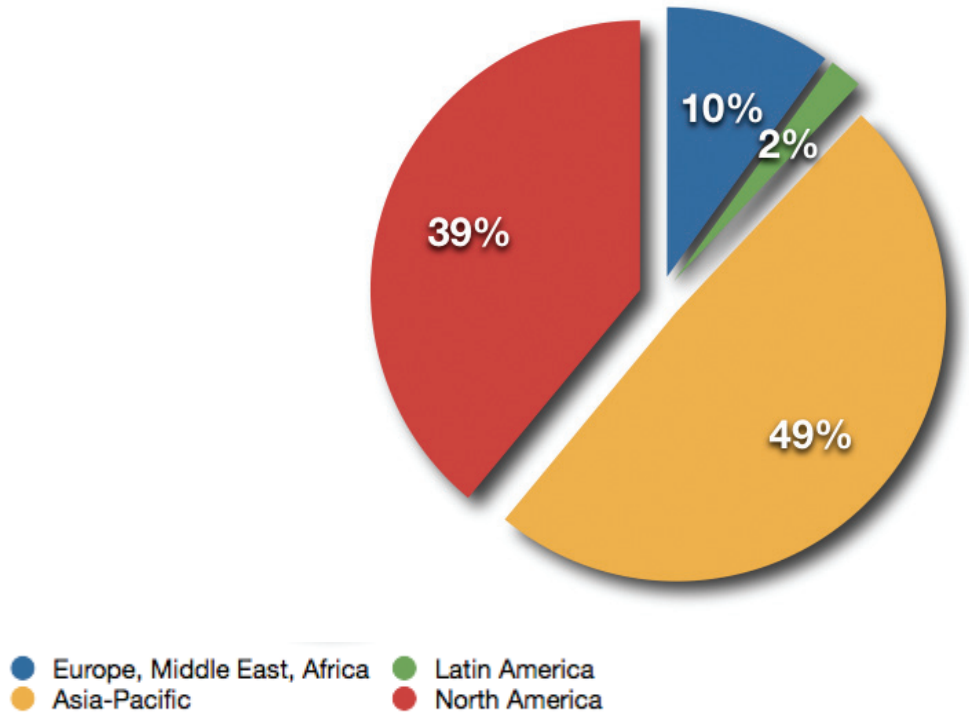
Nevertheless, global sales are expected to grow, especially in Asia where regional, language-specific devices can flourish. Following the trend towards format consolidation, a number of Japanese publishers and technology companies, including the Japan Electronic Publishers Association (JEPA), announced an intent to standardize on EPUB 3 [http://idpf.org/epub3_proposed_spec_released].

Japan represents a potentially lucrative market in digital reading, due to a large reading population and a tendency towards high-density living that does not support large physical book collections.

Ereader Revenue by Region, 2010



Ereader Revenue by Region, 2014



Source: Data from Yankee Group, 2010
[http://web.yankeegroup.com/rs/yankeegroup/images/2011-tablet-forecast-snapshot.pdf]

Unexpected outcomes

Sharp decline in retail print book business

By the first quarter of 2011, print sales in United States trade publishing had fallen 9% year-on-year, [http://www.publishersweekly.com/pw/by-topic/industry-news/financial-reporting/article/46891-print-stumbles-early-as-e-books-skyrocket.html] to 162 million units sold (based on 75% of unit sales both in physical stores and online, including the Kindle ecosystem). Consistent with all reported sales figures in 2010 and beyond, adult trade fiction has been most affected by the transition from print to digital; in some markets ebooks now outsell mass market paperbacks and even hardcover.

This trend has not yet reached non-fiction trade, with print sales declining only 1.1% in that segment according to BookScan. (Though juvenile nonfiction print *did* show a steeper decline at 11.7%, perhaps reflecting shrinking regional library acquisition budgets).

Trade non-fiction includes a number of publication types that have not yet been considered suitable for ebook conversion, including high-volume sellers such as cookbooks. It is to be expected that high-volume business guides and how-to books will make stronger entrants into the marketplace soon, especially as more advanced devices become commonplace. While the trajectory for academic publications will not exactly match the volumes demonstrated by trade publishing, the overall slope should be expected to start resembling the latter.

Substitutive versus additive sales

One question that immediately arises is whether an ebook sale is *substitutive* or *additive*. In a substitutive sale, a consumer buys an ebook rather than a printed book, and no further sale occurs. For price-sensitive buyers, this is obviously the typical model: an ebook is simply another format choice, analogous to choosing between a hardcover and a paperback. An additive sale can occur in one of two ways: a consumer can purchase *both* a digital and print version of the same title, generating two sales (or more often, choosing a print-plus-ebook bundle). A more common form of an additive sale would be when a consumer who *would have never purchased* the title in print chooses to buy an ebook edition, while print-centric purchases remain unaffected.

Amazon, citing increasing print sales co-occuring with skyrocking Kindle sales, believes that ebook effects

are additive [http://www.usatoday.com/printedition/news/20110105/1aebooksales05_st.art.htm], but the additive effect is more likely to be felt by a combined digital/print retailer like Amazon rather than an individual publisher. By capturing the loyalty of an ebook shopper, Amazon increases the likelihood that the same shopper will also turn to them for printed material (perhaps because an ebook edition is unavailable for that title).

Publishers and authors are more likely to view ebook sales as substitutive, which leads to considerable tension over differences in wholesale price — ebooks tend to be less expensive — and royalty rates — royalties paid to authors tend to be lower for an ebook sale versus a hardback sale.

The concern over substitutive sales could be ameliorated by leveraging the ability for digital books to reach untapped markets, particularly for English language titles overseas. Unfortunately, legacy geography-based regulations hamper the transition to global sales for all but the most forward-looking publishers. By delaying availability of English-language titles (or by providing them in translation only), publishers risk training overseas readers to seek out pirated editions rather than make legitimate purchases.

Academic publishers may be in a unique position to capture additive sales. With English as a near-universe language for the sciences and other disciplines, there is an urgent need to increase the availability of scholarly material in foreign markets. Targeted bundling of print *and* digital material may be of particular value in an academic context. Scholars, more than general readers, still prefer to have printed material which can be easily annotated. Scholarly marginalia is frequently more central to the activity of reading than the basic bookmarking or underlining that a pleasure reader may undertake.

While printed matter excels in such ad hoc illumination of texts, digital books are searchable in full text, can be analyzed for term frequency or subjected to other forms of meta-analysis, may be directly referencable through hyperlinks, can be read collaboratively by research partners, and may be enhanced by the availability of instant-lookup dictionaries and citation tools. Indeed, the ideal long-form scholarly publication is almost always one that is *both* digital and print-manifested; to date such bimodal material is not widely available, if at all.

Increasingly complex royalty/earnings

Traditionally, publishers operate under a "business-to-business" model. They engage directly with retailers,

not readers. Industry trade events are opportunities for publishers to connect with other publishers (typically to sell "secondary rights," such as translation or overseas distribution rights), to sell rights to purveyers of non-print formats (previously limited to audiobooks), or to promote titles to retailers and libraries. Though the largest such event in the industry, the Frankfurt Book Fair, does have some days open to the general public and feature direct book sales, most publishing events are limited to industry participants only.

This insularity is reflected in all aspects of the publishing financial ecosystem. Publisher sales forces work with booksellers and libraries, not consumers. Prices set by publishers frequently have little bearing on the final cost of a book that is purchased by a reader; it is expected that book sellers will deeply discount best-selling titles — many readers never pay full-price for new hardcovers, yet publishers and authors are compensated based on that price, the difference between absorbed by the retailers. The vestigal system of "returns" means that bookstores acquire books on consignment, rather than direct sales, and unsold books are returned to the publisher, at the publisher's expense. This is the traditional model of bookstores as wholesalers, rather than retailers who purchase inventory out-of-pocket.

In the ebook world, these rules are divorced from their historical precedents and no longer reflect reality. When Amazon engaged in aggressive discounting of ebooks to $9.99 or less, publishers reacted by wrestling control of pricing, resulting in the so-called "agency" model. It was feared that digital books, which are often released simultaneously with print hardcovers, would appear so radically underpriced that consumers would begin to expect these prices routinely. The existing publishing ecosystem is built on the model of selling many expensive (retailer-subsidized) bestsellers to fund a much larger pool of titles. Ebooks costing $9.99 or below threaten large publishing houses which depend on that pricing structure.

The outcome of publishers banding together to create the agency model (accused in some municpalities of being akin to price-fixing [http://www.independent.co.uk/news/business/news/oft-launches-inquiry-into-ebook-industry-2202570.html]) has been higher average prices across the board. Agency has negative effects on retailers; they are no longer empowered to use low prices as a method of attracting consumers, providing a tremendous advantage to entrenched players. Ironically this last effect may have been an unintended consequence, as publishers entered into the agency model specifically to benefit Apple as a new entrant in the ecosystem. With Amazon already having by far the largest customer base, there is less

incentive for customers to try a new service when price is consistently cited by readers as a motivator for shifting towards digital media.

Resurgence of interest in classic and out-of-copyright books

Nearly all ebook sites offer a mix of for-pay commercial content and out-of-copyright, free ebooks. Many devices come pre-loaded with perennial favorites such as *Pride and Prejudice* or *David Copperfield*. Nearly all such free ebooks (and many dubiously re-branded public domain ebooks put up for sale) are derived from Project Gutenberg, founded in 1971 by Michael Hart. Project Gutenberg, the work of thousands of volunteers, has much to teach about the capacity for collaborative transcribing, proofing, and formatting. (For a complete history of the project and its influence on the ebook revolution, see [http://www.etudes-francaises.net/dossiers/ebookEN.txt [http://www.etudes-francaises.net/dossiers/ebookEN.txt])

Scanned and computer-transcribed works are subject to errors (see Paul Duguid's now-classic excoriation [http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/1972/1847] of the quality of the texts in Google Books, but see also Dan Cohen's rebuttal [http://www.dancohen.org/2007/08/16/google-books-champagne-or-sour-grapes/]). One approach to correcting digital texts has been groups like Distributed Proofreaders [http://www.pgdp.net/c/], which has been correcting Project Gutenberg content for over a decade. The goal of Distributed Proofreaders is simple text accuracy, not scholarship. To that end it is worth watching the development of Oxford University Press's Oxford Scholarly Editions Online [http://oxfordscholarlyeditions.com/], which will aim to present definitive texts of material and editorial comment.

Scramble to allocate rights from out-of-print/pre-digital works

Publishing contracts signed before the proliferation of digital media obviously did not include specific clauses about ebook rights, and many older contracts specified exactly which rights were conferred to publishers from authors (such "United States hardback editions"). This has led to a state where many valuable in-copyright works by known famous authors do not have clear rights assignments for digital content.

In one highly-publicized digital venture, the controversial literary agency Wylie Agency spun out Odyssey Editions [http://www.odysseyeditions.com/] in July 2010, negotiating an exclusive contract with Amazon to sell digital versions of books still in print with traditional publishers, including Random House and Penguin.

Publishers responded negatively [http://www.guardian.co.uk/books/2010/jul/23/publishers-wylies-ebook-deal-amazon], and within a matter of weeks titles claimed by Random House were removed from the Odyssey catalog [http://latimesblogs.latimes.com/jacketcopy/2010/08/agent-amazon-partnership-ebooks-random-house.html]. Other titles remain for sale under the existing terms, but no further books have been released at the time of this writing.

The public scuffle between Random House and Wylie Agency did not end the rights disputes. Open Road Integrated Media [http://www.openroadmedia.com/] publishes dozens of ebooks that were previously released in paper editions by transitional publishers. Recently, Amazon has been asserting itself as a publisher, hiring a well-known publishing executive and literary agent to head its New York-based publishing wing [http://www.thebookseller.com/news/amazoncom-hires-larry-kirshbaum.html]. Significantly, Amazon does not appear to intend to use its imprints to publish new authors but instead draw already-successful authors away from their houses through more favorable terms. Amazon's five imprints [http://www.techflash.com/seattle/2011/06/amazon-ramps-up-publishing.html] are not limited to digital distribution alone; they will compete for shelf space with other publishers' titles in brick and mortar stores as well as enjoy prominent promotion on Amazon.com [http://www.amazon.com/].

Authors themselves are increasingly choosing to reclaim their own digital rights independently. Releasing out-of-print works can be of critical importance to series fiction writers [http://www.teleread.com/paul-biba/an-interview-with-self-publishing-author-patricia-ryan/] in particular, who may still have later series books in print while earlier books are unavailable. Re-publishing backlist series digitally means that authors can experiment with low or even free pricing of early books in the series as a draw for readers to purchase later titles.

Migration of authors towards self-publishing or digital-first publishing

In addition to those authors who reclaim expired or undefined digital rights, many authors are foregoing traditional print publication entirely. Some authors, like Amanda Hocking, [http://www.nytimes.com/2011/06/19/magazine/amanda-hocking-storyseller.html?_r=1&pagewanted=all] chose digital/self-publishing after rejection by traditional publishers (and notably has since signed with St. Martin's Press). J.A. Konrath [http://www.jakonrath.com/] is successful in both fiction publishing and in assisting other authors in their self-publishing efforts. Others, most notably J.K. Rowling, withheld digital rights deliberately and chose

to release ebook editions directly [http://www.pottermore.com/].

Hocking used the self-publishing service Smashwords [http://www.smashwords.com/] as her distribution channel; to date there is no comparable service for self- or digital-only publishing in an academic context.

Self-publishing in academics

To date, there are no significant examples of academic self-publishing outside of dubious "author mills." There is evidence that some digital-first publishing houses have been soliciting theses and other publications from scholars [http://chronicle.com/forums/index.php?ChronicleUser=c80atumlqe5puc3m15oft0e9m4&/topic,45997.0.html] or accepting computer-generated papers for publication under a supposedly peer-reviewed system [http://scholarlykitchen.sspnet.org/2009/06/10/nonsense-for-dollars/]. That otherwise accomplished scholars sometimes "fall" for low- or no-quality publication houses suggests that there is unmet demand in providing digital outlets for academic publications that don't neatly fall into the journal article/monograph/textbook categories.

Use of blogging software

Originally designed for short-form journal entries, blogging platforms have received a great deal of attention as viable authoring environments for longer-form or structured work. Today, "blogging software" no longer principly means software used to provide individual web journal articles, but instead can apply to any kind of writing.

Although most self-published authors use traditional word-processing tools, there is an increasing trend of "blog to book" deals. While such books tend to be emphemeral and non-studious in nature, it is clear that there is benefit to platforms that support both authorship, editing, commenting, and possibly direct to digital book publication.

TABLE 4. ONLINE PLATFORMS USED FOR PROFESSIONAL AUTHORING/DRAFTING

Online platform	Specialty	Use in publications
Tumblr [https://www.tumblr.com/]	Image-heavy short form content	Numerous books, largely humor. [http://www.tumblr.com/spotlight/book+deals]
MediaWiki [http://www.mediawiki.org/wiki/MediaWiki]	Collaborative editing	Wikipedia, features output paths to PDF and other formats
Papers [http://www.mekentosj.com/papers/]	Research gathering	Manuscript production
Mandeley [http://www.mendeley.com/]	Research gathering and journal search	Reference management and export
Scholarpedia [http://www.scholarpedia.org/article/Main_Page]	Peer-reviewed academic articles	Authors become editors on publication

Also see (in Appendices) an expanded chart of specialized ebook/authoring platforms.

Open-source textbooks

One activity that is seeing an increasing rise in demand for educational publications has been in the competitve space of digital textbooks. Most digital textbooks are currently produced by traditional publishers and delivered on platforms such as VitalSource Bookshelf [http://www.ingramcontent.com/MRKNG/2011/52373/52373Vitalsource.html] or upstart for-profit technology companies such as Inkling [http://www.inkling.com/]. These platforms offer interactive and multimedia features, attempting to appeal simultaneously to students looking for a livelier reading experience, and to publishers who feel that "enhanced" textbooks are a deterrent to piracy or used print textbooks.

On the other end of scale are ventures to produce inexpensive, even free, etextbooks. Some are truly non-profit, such as CK-12 [http://www.ck12.org/flexbook/] or the Open University [http://www3.open.ac.uk/media/fullstory.aspx?id=19962], which offers interactive textbooks generated with an open-source XML workflow[http://sourceforge.net/projects/bookbind/]. Others such as Flat World Knowledge [http://www.flatworldknowledge.com/] are for-profit companies which publish textbooks under a Creative Commons [http://creativecommons.org/] license and shares revenue with authors. Flat World and CK-12 both use hosted WYSIWYG editors for etextbook creation.

Though digital textbooks have been produced using a variety of platforms, from XML workflows to Wikipedia-like collaborative editing tools, it is likely that many, if not most, etextbooks in the next few years will use EPUB 3 or HTML5, including those from Inkling [http://venturebeat.com/2011/08/04/inkling-17m-digital-textbooks/],VitalSource [http://www.mediabistro.com/ebooknewser/macmillian-and-ingram-adopt-epub-3_b11263], and Chegg [http://techcrunch.com/2011/08/18/textbook-rental-giant-chegg-goes-digital-now-offers-html5-books/]. Key to that technology choice is the support for advanced mathematic layout and interactivity that is afforded by those standards.

Perceived value of a digital book

According to the Book Industry Study Group, price and availability of free samples are the most influential factors leading to an ebook purchase. (BISG, 2011) [http://www.bisg.org/news-5-631-press-releasebisg-study-reveals-e-book-buyers-are-accelerating-their-move-away-from-print.php]. Publishers blame retailers for setting the perception that ebooks are intrinsicallly inexpensive, but consumers naturally feel that a physical object should cost more than a digital representation of the same content.

Consumers also tend to believe that digital books *should* cost significantly less than a printed book because there is no cost for printing, shipping, and warehousing. In fact, artifacts of the print process account for only about 7.5% of the cost of a retail hardback book [http://www.newyorker.com/reporting/2010/04/26/100426fa_fact_auletta]. However, consumers are correct that ebooks have a number of factors which ought to reduce the cost:

1. Ebooks are often produced from masters which have not gone through the complete editorial cycle and may still contain errors.
2. Errors may be introduced post-hoc due to ebook conversion processes or inadequate training.
3. Digital rights management and other rights restrictions prevent consumers from re-selling, lending, or transferring their purchases.
4. Many ebook systems do not permit highlighting or annotations, and disable printing to apply hand-written annotations to paper fascimilies.
5. Ebook retailers can and do close shop, stranding purchases on ecosystems that are no longer maintained.

Publishers counter that ebooks often a number of features that are lacking in their print counterparts,

including adjustable font sizes, automatic bookmarking, and that some titles benefit from multimedia enhancements that would not be possible in print. Curiously, the absence of a physical artifact can be seen either as a feature or a deficiency, depending on the type of book and the storage capacity of a reader's home.

Frustration with specification limitations

EPUB fully became the de facto ebook standard within two years of its release, a success for a specification by any measure. (Even Amazon, which does not use EPUB as a distribution format to its users. does use it as an input format from publishers.) Publishers of text-heavy books, such as novels, monographs, and some non-fiction, found EPUB sufficient for their needs, and the nature of monochrome electronic ink screens meant that books with extensive photographic or color content was not a candidate for the first ebook releases.

With the advent of tablets and high-resolution mobile screens, publishers of textbooks, cookbooks, children's fictions and other design-heavy works began to demand accommodation in the marketplace. Responding to a need for richer design options, Apple moved to extend the EPUB format to include a variant dubbed "fixed-layout EPUB" [http://code.google.com/p/epub-revision/wiki/iBooksFixedLayout]; Barnes and Noble released a similiar extension for their nook platform designed to accommodate magazine content, called Digital Replica Plus [http://code.google.com/p/epub-revision/wiki/BNFixedFormat].

In both cases, these "higher fidelity" formats sacrifice some of the advantages of the original EPUB specification — their adaptability to any screen size or capability — by fixing the viewport at a pre-defined aspect ratio and scaling the content to accommodate. This scaling means that much of the content designed for these formats can only be effectively viewed in tablets or larger screens. In many ways this approach provides no additional benefits to merely reproducing print-ready PDFs; indeed many "fixed-layout" books produced by publishers to date are simply images of print pages. Producing such content can be expensive [http://ebookarchitects.com/conversions/fixedlayout.php], and it is unclear if pricing can be adjusted upwards to recoup the investment.

Fixed-layout EPUB does potentially allow for features that are difficult or impossible to do with PDF alone, including multimedia and interactivity. Thus far, though, the variant has been more of a publisher-driven

enterprise, meant to meet the intermediary goals of an industry still shifting away from print.

EPUB 3

To avoid further fragmentation in the service of high-design publications, a revision to the initial EPUB specification (called EPUB 2 for obscure reasons) was fast-tracked; EPUB 3 was released in October 2011. EPUB 3 includes a number of extensions to the original EPUB 2 specification, particularly related to design and layout:

1. Explicit adoption of the richer layout capabilities of the latest web standards, HTML5 and CSS 3.
2. Extensions to allow for multimedia books, including hybrid audio/print books.
3. A modular extensibility mechanism, both at the specification/technical layer and at the committee level.

An unstated goal of EPUB 3 was to prevent the necessity of "EPUB 4" within a short time frame. The extension mechanism added to the IDPF process is meant to accommodate more frequent updates to EPUB 3; the first of such initiatives to formalize is the attempt to harmonize the existing fixed-layout formats. While the initiative is being driven by distinctly non-academic publishers, textbook authors have been following the work with interest. Additionally the work coming from Barnes and Noble on coordinating different "renditions" of the same work (print, fixed-layout, reflowable) are of potential scholarly interest as it develops. The subcommittee on EPUB Fixed-Layout met in Taiwan in October 2011 and continues to explore rendition mapping [http://code.google.com/p/epub-revision/wiki/RenditionMappingAdHocGroup] and related metadata concerns [http://code.google.com/p/epub-revision/wiki/RenditionMetadataAdHocGroup].

Dictionaries and indexes

Other features deferred from the initial release of EPUB 3 are also of interest to academic audiences:

1. Extensions to the core EPUB vocabulary to allow for publication of rich reference works, particularly Dictionaries [http://code.google.com/p/epub-revision/wiki/DictionariesCharterProposal]. This work also covers allowing ebooks to be bundled with custom, domain-specific dictionaries, as well as for multi-lingual dictionaries.

2. The Indexes charter proposal [http://code.google.com/p/epub-revision/wiki/IndexesCharterProposal] would seek to address the incomplete specifications for indexing in EPUB 3 content. Items in scope include indexes which may include definitions in-context (such as popups), cross-publication indexes, and indexes which can be explored either in forward or reverse order (finding index terms in content versus finding content from index terms).

Hardware considerations

After a series of early failures (Rocketbook, 2000), the Kindle device demonstrated that hardware designed almost exclusively for reading ebooks was economically viable. Though due in large part to the availability of a large pool of highly-desirable content, part of the success of the device and its successors have been due to its screen technology, known as electronic ink. Sony was the first major company to produce professional-grade e-ink hardware, but lacked the wireless connectivity and deep store catalog to attract more than a niche group of early adopters. In 2011, major book retailers such as Barnes and Nobile and Kobo have their own branded devices, which continue to sell briskly and be incrementally improved.

Electronic ink has a number of characteristics that make it ideal for consuming textual content. It has good contrast (although the e-ink light color is light gray rather than true white). Measured in dots-per-inch, e-ink display resolution is matched only by a few best-of-breed liquid crystal displays (LCD), such as the iPhone 4 "retina" display. Most importantly, e-ink is "bistable," meaning that it requires no power usage to remain in the "on" state. This feature, and the lack of a backlight, means that electronic ink devices have battery life ratings measured in weeks or even months; by contrast, even highly-efficient phones and tablet devices have battery lives measured in hours when under continuous use.

Text clarity and battery longevity come with a few drawbacks. Most notably, true color e-ink displays do not yet exist; even prototype devices have only lackluster color rendering. Electronic ink screens have a noticable delay when refreshing, though this limitation is visibly improved in each screen generation. Lack of color and long refresh together mean that video is impractical, though the devices can and do play audio.

The marketplace has dealt with these limitations by diversifying offerings. Amazon and other retailers offer ereading applications for mobile phones, computers, and tablets. Barnes and Noble offers a version of the nook that uses a traditional color LCD display, marketing it explicitly towards readers with children and/

or an interest in color magazines or cookbooks. Since many owners of specialized e-ink devices also own smartphones or tablets, the need for a discrete color-capable ereader is minimal.

Form factor plays a roll in how ebook content is most comfortably consumed. Hardware eink devices are typically shaped like trade paperback books, with a width/height ratio of 3:4. This is ideal for textual content, providing a comfortable line length of approximately 40 characters per line with default settings. However, this length is insufficient for many charts, tables, and graphs, a problem which becomes acute when considering scholarly content. Thus activity around digital scholarly publishing has traditionally been focused on the desktop computer (where screen size is fluid and relatively unbounded), and more recently tablets, which provide a similar 3:4 ratio in portrait mode but larger displays than most ereaders. Additionally tablet display technology makes zooming or panning across large figures more user-friendly; these activities are slow and awkward on electronic ink screens.

Despite challenges with page fidelity, the benefits of digital reading on smaller mobile devices should not be overlooked. While not every aspect of every piece of content is appropriate for reading on a phone-sized screen, readers make appreciate the ability to consume abstracts or excerpts of longer pieces in a mobile context, provided they can easily mark those passages for close reading in other modalities. Also worth considering is that screen technology improves year-on-year; text on a high-resolution phone display is readable at much smaller sizes than on devices only a few years old. More important, though, is the fact that mobile devices are ubiquitous, personal, and always at hand; even if they are not the locus of sustained content consumption, they are a critical adjunct.

ACADEMIC CONSIDERATIONS

ACADEMIC CONSIDERATIONS

Although this report largely uses the specific term "book" or the broad term "publication", there are distinct types of scholarly objects that can be published digitally. These objects have different format and distribution implications, as shown:

TABLE 5. SCHOLARLY PUBLICATION TYPES AND APPLICATION FORMATS

Scholarly publication	Digital format options	Static or fluid?	Authoring implications
Anthology	PDF, EPUB	Either	Many, with centralized editing
Monograph	PDF, EPUB	Static	Single author, formal publication
Single chapter	PDF, EPUB, HTML, blog post	Static	Single author, may be recontextualized or remixed
Textbook	PDF, EPUB 3, HTML5	Either	Possibly many authors, possibly containing enhanced content
Journal article	PDF, EPUB, HTML, blog post, database entry	Static	May imply commenting
Journal letter	HTML, blog comment	Either	May be "published", or simply a formalized comment
Editorial	PDF, HTML	Static	
Working paper	PDF, EPUB, HTML, blog post	Fluid	May imply commenting
Conference proceeding	PDF, EPUB, HTML, database entry	Either	May be fixed, or may be part of "living" conference site
Theses	PDF, EPUB	Static	May be fluid during authoring process

Existing academic ebook platforms

Ebook distribution platforms have existed long before the Kindle "revolution", providing digital book content as subscription databases to research libraries and educational institutions. It is useful to distinguish

between these ebook platforms—typically based around PDF and often accessible only through computers with active Internet connections—with the more contemporary use of ebooks—as discrete files readable on a range of devices. While a complete assessment of all extant ebook platforms is outside the scope of this paper, a small subset is described for context.

Cambridge Collections Online

The Cambridge Collections Online offering is typical of most ebook collections in that the titles are in PDF form, with limited HTML-based delivery of summaries and excerpts for search purposes. While citations may be downloaded for use in citation management software, bookmarks and annotations are restricted to the platform itself and not integrated into the content or extractable. PDFs are available for download, with printing/copy-paste restrictions applied to the files as specified by the publisher. The ebook platform is provided by Semantico [http://www.semantico.com/], used by a number of publishers including Blackwell Reference Online [http://www.blackwellreference.com/subscriber/uid=2748/?authstatuscode=202].

Springer eBooks

The Springer eBooks collection includes monographs, contributed volumes and conference proceedings as well as textbooks and major reference works. The collection is primarily PDF but does include some HTML content.

Most science, technical, and medical content (STM) is produced in NLM XML, which is easily convertable to HTML/EPUB. Similarly, content in TEI or other humanities-related XML formats are also convertable to EPUB using open source tools. Nevertheless the majority of online ebook collections are based on PDF or other "page-fidelity" systems.

The Springer eBook collection is unique in that it does not include DRM on the downloadable ebooks. The ebooks can be transferred to any number of devices and reading products, allowing users to consolidate content onto their preferred system and providing for access on multiple devices simulatneously. EPUB is not offered as a downloadable format; only PDF is available. This limits the range of suitable devices to large-screen tablets and ereaders and standalone reading systems on computers. A subset of titles are available in HTML, which would provide the highest level of accessiblity and reflowability, but that HTML is not downloadable nor is it packaged in EPUB form.

VitalSource

Ingram VitalSource [http://www.ingramcontent.com/MRKNG/2011/52373/52373Vitalsource.html] provides textbook content to students on a per-site basis, often as the primary provider for individual institutions. VitalSource offers rich interactive content for some titles, but the majority are "page-fidelity" representations of textbook PDFs. In 2011, VitalSource announced support for EPUB 3 [http://www.econtentmag.com/Articles/News/News-Item/Ingram-Announces-EPUB-3-Support-for-VitalSource-75665.htm] and has indicated to its publishing partners that its preferred input format in 2012 and beyond will be EPUB 3 and not its proprietary XML format. Other textbook platforms such as CourseSmart have similarly suggested that EPUB 3 will be part of their long-term content delivery strategy [http://idpf.org/epub3_proposed_spec_released].

Authoring content

PDF-based authoring is generally a simple matter of "Save as PDF" with limited need for quality control or assurance. As PDF is a "page fidelity", viewport-based format, the output from a PDF workflow is rarely unexpected. Scholars wishing to engage in reflowable digital publishing (using HTML or EPUB) will need to work with tools that can approach the simplicity of saving as PDF or printing pages. To date, no tools exist that are exactly that simple, though blog-based tools (see Appendices) or (for the adventurous) text-markup-based tools can approach a level of reliability that becomes feasible for serious work.

Tabular content

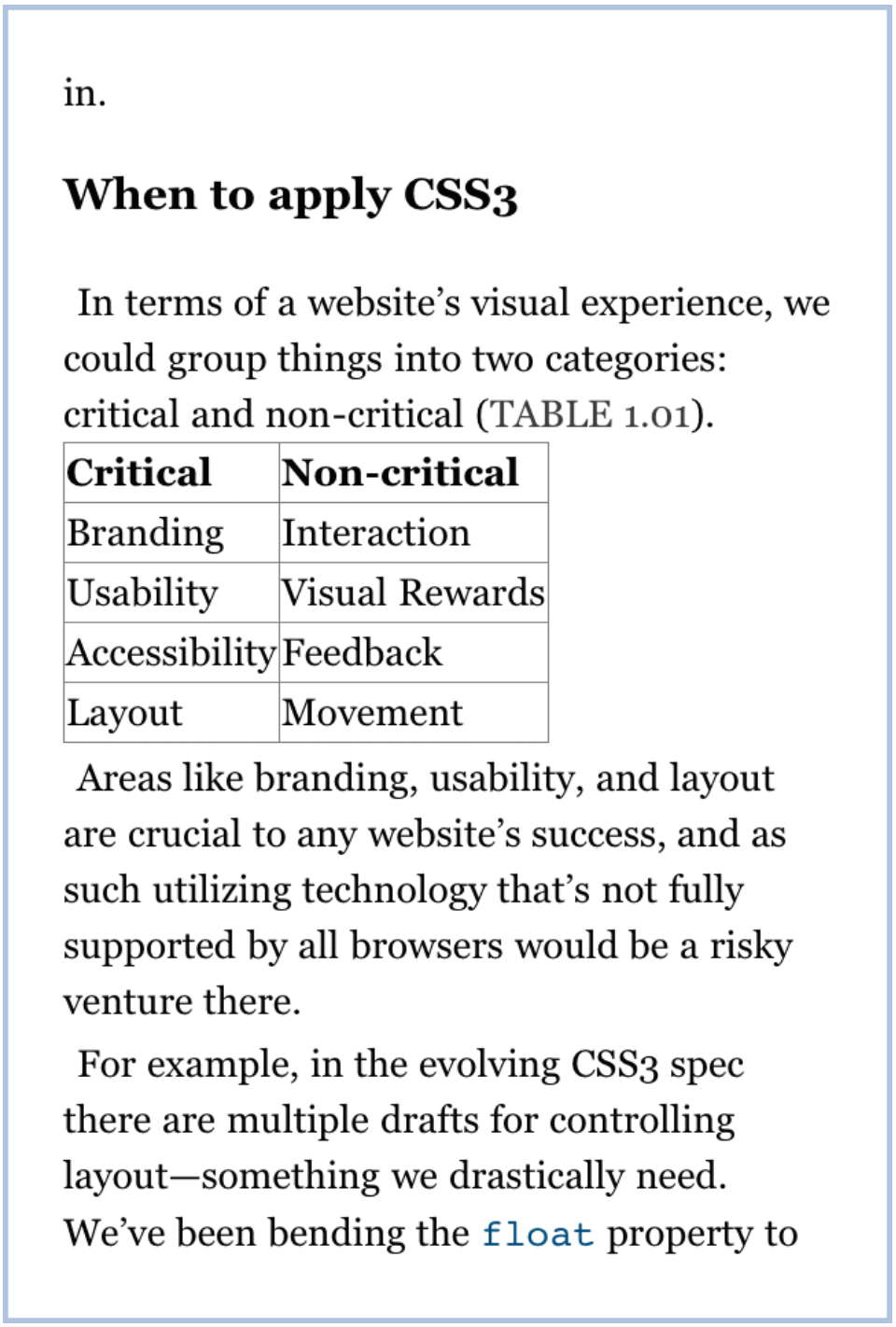
While tables, even arbitrarily-complex ones, are supported in all major ebook formats, tabular data can be problematic for ereading devices with narrow viewports. Phone-sized screens, in particular, rarely render tables in a legible fashion. Ebooks that may have a high percentage of mobile reading activities should generally avoid tabular data (which can easily exceed the horizontal width of the viewport) in favor of alternate data presentation forms that do not rely on a large number of horizontal columns.

Large or wide tables are frequently cut off on narrow screens

into advanced syntax and real-world usage later.

Property	Supported in		
border-radius			
	3+	3+	1+
text-shadow			
	1.1+	2+	3.1+
box-shadow			
	3+	3+	3.5+
Multiple background images			
	1.3+	2+	3.6+

Narrower tables scale well on many ereading devices



in.

When to apply CSS3

In terms of a website's visual experience, we could group things into two categories: critical and non-critical (TABLE 1.01).

Critical	Non-critical
Branding	Interaction
Usability	Visual Rewards
Accessibility	Feedback
Layout	Movement

Areas like branding, usability, and layout are crucial to any website's success, and as such utilizing technology that's not fully supported by all browsers would be a risky venture there.

For example, in the evolving CSS3 spec there are multiple drafts for controlling layout—something we drastically need. We've been bending the `float` property to

Academic Considerations

Screenshots from CSS3 for Web Designers by Dan Cederholm [http://www.abookapart.com/products/css3-for-web-designers]

Mathematical notation

Science, technical, and medical content typically features mathematical notation. There are two leading ways to represent math in textual markup:

1. TeX [http://en.wikipedia.org/wiki/TeX]

2. MathML [http://www.w3.org/TR/MathML/Overview.html]

Neither markup language was directly supported in the EPUB 2 specification. Although mathematics can be rendered in a publication as a graphic, doing so has a number of negative consequences:

1. The formula will not scale up or down in font size
2. The formula may not wrap appropriately in a constrained viewport
3. When rendered as an image, the graphic becomes "invisible" to screen-readers and other assistive technologies

In EPUB 2, the preferred solution to creating scalable (but not accessible) mathematical notation was to use Scalable Vector Graphics (SVG), a method for drawing lines in a digital document.

Mathematical formulas rendered using SVG and EPUB 2

$$\text{Sbit} = \frac{\text{Total\_number}^{\text{Packets}}\text{\_sent}}{\text{Simulation\_Time}} \text{ [Mbps].} \quad (6)$$

The sending bit rate or throughput *Jain's fairness index* is estimated according to the following equation:

$$f(x) = \frac{\left(\sum_{i=1}^{n} \alpha_i\right)^2}{n\left(\sum_{i=1}^{n} \alpha_i^2\right)} \quad \text{where } \alpha_i \geq 0, \quad (7)$$

where n is the number the contending flows, and α is *sending bit rate* (Sbit) or *throughput* (Thr). If all flows get the same amount of α (sending bit rate or throughput), then the fairness index equals 1, thus the network is 100% fair [22].

Formulas rendered this way can be visually attractive, however they suffer from the same "invisibility" problem as mathematics captured as raster images.

EPUB 3 has added presentational MathML support [http://idpf.org/epub/30/spec/epub30-contentdocs.html#sec-xhtml-mathml], which greatly increases the reach of scientific markup possible in digital books.

Mathematical formulas rendered using MathML and EPUB 3

Additionally, the equation above tells us that every eigenvector of S for ρ_i is again an eigenvector of A for ρ_i^2. So for $1 \leq i \leq q$, the relevant eigenspaces are related by

$$\mathscr{E}_S\left(\sqrt{\lambda_i}\right) = \mathscr{E}_S\left(\rho_i\right) \subseteq \mathscr{E}_A\left(\rho_i^2\right) = \mathscr{E}_A\left(\lambda_i\right)$$

Presentational MathML allows screenreaders or other assistive technologies to accurately read and translate the markup using text-to-speech. A sophisticated MathML rendering system could also know how to line-wrap mathematical figures, providing a richer and more flexible layout.

Note

EPUB 3's support of MathML is purely *presentational*. The more complex subset of the MathML specification known as Content Markup [http://www.w3.org/TR/MathML/chapter4.html#contm.intro] represents the underlying mathematical meaning of the formulae. In a compatible rendering system, Content Markup could allow a system to evaluate the formula. This is currently out of scope for ebook rendering systems.

LaTeX in digital books

Many scientists are accustomed to writing math in TeX or LaTeX rather than equation-editing tools that can output MathML (it would be unusual to author directly in MathML). Math on the web is frequently rendered by the open-source tool MathJax; it accepts both LaTeX and MathML input forms. MathJax is compatible with scripting-capable ereading systems and may be an attractive interim solution to displaying mathematical content either when only LaTeX content is available, or for systems which do not yet support MathML natively.

Fundamentally, though, TeX is a viewport-aware technology which is not syntactically or even philosophically aligned with XML-based formats, and all current ebook formats are XML-based. A more forward-looking

workflow may be to allow LaTeX math as an input markup format, but to deliver any math-heavy ebooks using MathML, mechanically translated from LaTeX.

Task-driven activities

Scholarly engagement with publications comes in distinct activities which have better or worse affordances in the digital realm at the moment:

1. Reading
2. Searching
3. Citing and linking
4. Annotating
5. Portability

Reading

In the autumn of 2009, seven American universities participated in a pilot program investigating the usage of Kindle DX eink devices. Unlike most other eink devices, the Kindle DX model is large enough to view US letter/A4 PDF documents at close to their normal paper size (with 9.7-inch diagonal screen size).

Results from the pilot were mixed. Princeton University reported some positive findings [http://www.princeton.edu/ereaderpilot/eReaderFinalReportLong.pdf], though in the final analysis concluded:

> In short, students were very positive about the "reading" aspects of the present crop of e-readers when compared to paper, but felt that the experience of writing and studying could by vastly improved by better input tools, by having color available for highlights (but not necessarily for the page being read), and by allowing a way to skim visually or flip through the readings, in a way that was more analogous to using paper.

Students were asked to rank the most-desired improvements:

1. The ability to highlight and annotate PDF files
2. Improving the annotation tools
3. Providing a folder structure to keep similar readings together
4. Improving the highlighting function
5. Improving the navigation within and between Kindle documents

Since the time of the pilot some of these features have been introduced into the Kindle system, including collections. Students also cited the slow pagination as a deficiency of the device; as this is related largely to the hardware technology, improvements will have to come with new devices rather than improved software.

It is worth noting that the students particularly did not request additional "features" beyond improved ability to annotate and skim/browse through documents. Participants in the pilot felt that an ereader whose feature set was set apart from a general purpose computer had value. "The perfect reading tool would […] simply be like a conventional book, but paperless. I would prefer not even noticing that I am using an electronic device to read."

In a report published in *Proceedings of the 2011 annual conference on Human factors in computing systems*, researchers from the University of Washington [http://www.princeton.edu/ereaderpilot/eReaderFinalReportLong.pdf] reported that:

Note

> "…although the Kindle DX was poor for responsive reading, which typically involves marking up, they found that it was still acceptable for receptive reading situations.

Searching

Search features on existing ereading devices, when they exist, are frequently rudimentary. While it appears that the Kindle does support stemming [http://www.mobileread.com/forums/showthread.php?t=116458] in its built-in search capability, there is no ability for users to search across their collections in any existing ebook readers. While Google Books [http://books.google.com/books] does have rich search capabilities across their corpus, is not integrated with their ereader software, and it cannot be limited to the user's personal library.

Scholars do have access to a rich set of search tools on library database systems, but for those collecting digital books for their own use, there is currently no rich full-text search with a feature set like Solr [http://lucene.apache.org/solr/] or similar products. Instead, search is primarily a mechanism for initial discovery; in a study conducted by Springer at the University of Liverpool [http://www.springer.com/cda/content/document/cda_downloaddocument/V7671+Liverpool+White+Paper+Part

2.pdf?SGWID=0-0-45-1037538-0], a majority of students (73%) and faculty (67%) performed keyword searches on selected topics to locate ebook content versus other acquisition strategies.

Browsing

Academic "browsing" behavior can be considered a subset of searching behavior, where the inquiry is usually topic-based rather than an attempt to discover a particular paper, author, or fact. Online tools facilitate browsing by providing high-level access to a large range of content across different resources.

To date, ereading devices and software do not facilitate browsing across a multiplicity of titles other than providing lists of basic metadata in the user's collection, and (sometimes) the ability to organize ebooks into user-defined categories or folders. Even the highly-browsable online stores as provided by Amazon or Barnes and Noble do not make it easy for users to quickly browse book content. Yet "power browsing" — skimming a range of content to identify which ones to spend time on in-depth — is a common research activity.

Ereading services could better support this behavior by allowing researchers to harvest large pools of content for use while traveling or otherwise away from the research institute, and enabling those users to mark content of interest for full consumption at a later time.

> Students power browse rather than read online. An e-book page can be scan read in about 1 minute. Only 5% of users spent more than 5 minutes viewing a page and 85% spent less than 1 minute viewing a page. This indicates that a large amount of power browsing and downloading is occurring.
>
> "Dispelling myths about e-books with empirical evidence", JISC Collections and CIBER report [http://www.ucl.ac.uk/infostudies/research/ciber/downloads/JISC-summary.pdf]

It is worth noting that the Springer eBook study found that the "preferred method for locating material was keyword search" rather than browse [http://www.springer.com/cda/content/document/cda_downloaddocument/V7671+Liverpool+White+Paper+Part2.pdf?SGWID=0-0-45-1037538-0].

Citing and linking

The mechanisms used by scholars to cite digital documents are out of the scope of this report, but even informal linking to ebooks is a problem domain recognized by the larger digital book community as a challenge.

EPUB 3 defines a method of "deep-linking" into a .epub file to a chapter, section, or even arbitrary set of words. The EPUB Canonical Fragment Identifier [http://idpf.org/epub/linking/cfi/epub-cfi.html] specification allows a document author to refer to any content inside an EPUB publication. As of this writing, EPUBCFI is a proposed specification that has not yet been adopted by any publicly-available reading systems. Additionally the notation is highly complex and meant to be authored by automatic tools rather than human indexers or document authors.

EPUBCFI only solves one half of the citation/linking equation; it identifies a position inside a given ebook file. It does not purport to identify the ebook itself, or where it might be found on the internet or in a particular content repository. Although a working group inside the IDPF [http://code.google.com/p/epub-revision/wiki/LinkingImplementationProposals] was developed to facilitate this work, consensus could not be reached in time for the EPUB 3 specification deadline. That work has not yet resumed in the post-release cycle for EPUB 3 revisions.

Annotating

Linking is a key component in other topics such as social reading and shared/portable ebook annotations. To that end, in July 2011 the Internet Archive and National Information Standards Organization (NISO) announced a meeting, funded by the Mellon Foundation, to address formally specifying a linking and annotation syntax. Meetings will take place in both Frankfurt and San Francisco [http://www.niso.org/topics/ccm/e-book_annotation/] in October 2011.

An additional limitation faced by scholarly users is that existing ereader systems for annotations and note-taking are not designed with their needs in mind. Note-taking mechanisms are often awkward (suitable for only brief notes or simple highlighting), and publisher- or retailer-enforced limitations on number of highlights or notes may conflict with scholarly goals.

> Although the Kindle DX supports a rudimentary annotation system, the ability to superimpose or juxtapose marks on texts is crucial for responsive reading, but also for other types of reading that may have different goals [than purely receptive reading].
>
> *Proceedings of the 2011 annual conference on Human factors in computing systems*
> — http://portal.acm.org/citation.cfm?id=1979375

Portability

It is clear that ebook readers do not provide a user experience that beats working in an office environment. However, during focus group sessions carried out by the jiscPUB thinktank it was noted that the enhanced portability over desktop and even laptop computers do enable readers to work in unexpected places.

"Really invaluable for fieldwork; the ability to access all your texts while in the field in Africa is fantastic. Currently I use a laptop, but it is not so portable and the battery life is not so good. In some places they only have one hour of electricity a day so it's hard to keep everything charged."

—Early Career Researcher, School of Social and Political Science, University of Edinburgh

The focus groups also indicated that ereaders offer opportunities for researchers to change and enhance working practices that were not previously possible.

"The ability to have my research data all in one place is great – I have digitised all 50 or so of my primary texts and added them to the GoodReader library on the iPad. It means when I travel to conferences I don't have to carry the books or laptop around, so there's no risk of damage to the originals, and I can work on the go."

—Postgraduate Student, School of Social and Political Science, University of Edinburgh

Accessibility

Providing highly-accessible devices and software is critical to any use of ereading software in an educational setting. Early pilots with the Kindle DX device faced legal challenges [http://www.libraryjournal.com/article/CA6716860.html] in the United States because the devices did not have full text-to-speech capability (the accessibility of the Kindle platform was significantly improved in 2010 [http://serotalk.com/2010/09/28/an-accessibility-review-of-the-amazon-kindle-3/]). Ereading software on third-party platforms, such as iOS, can use the built-in accessibility features [http://www.smh.com.au/digital-life/computers/blogs/gadgets-on-the-go/voiceover-and-ibooks-on-the-ipad-20110211-1ap4r.html], but the ability for library users to transfer content to such devices may be hampered by digital rights management applied to the content. For this reason, some collections, such as Springer eBooks, specifically cite an absence of DRM as a feature: "Users can choose how they wish to access content. Springer eBooks can be delivered and utilized

on the widest array of devices." (Springer eBooks website [http://www.springer.com/librarians/e-content/ebooks?SGWID=0-40791-0-0-0]).

The EPUB specification itself has support for accessibility features, including Digital Talking Book content [http://idpf.org/epub/20/spec/OPS_2.0.1_draft.htm#Section2.4] and the Navigation Center [http://idpf.org/epub/20/spec/OPF_2.0.1_draft.htm#Section2.4.1] file. The more recent EPUB 3 specification further extends accessibility to include rich semantics, improved navigation, and synced audio/text [http://idpf.org/epub/30/spec/epub30-overview.html#sec-accessibility]. Providing highly-rich semantics and navigation must be done at the point of authoring, but typically most valid EPUB 2 documents are acceptably accessible, and a major improvement over the accessibility of most PDF documents.

RECOMMENDATIONS FOR FUTURE WORK

RECOMMENDATIONS FOR FUTURE WORK

1. Provide rich search tools for individual collections of ebooks

Open-source full-text search tools such as Solr [http://lucene.apache.org/solr/] are in wide use in scholarly repositories and research settings, but not found in private ebook collections. One piece missing from a scholarly-centric ereading system is the ability to search across those works using language-dependent stemming, capable of leveraging the presence of semantically-rich XML in some ebook formats such as EPUB 3.

2. Tools for generating or traversing ebook citations

There are numerous online citation systems, the most popular being Zotero [http://www.zotero.org/]. However, Zotero does not provide a method to cite EPUB publications, especially not selections of text inside of ebooks. The EPUB 3 standard includes a linking component known as EPUB Canonical Fragment Identifier [http://idpf.org/epub/linking/cfi/epub-cfi.html] (EPUBCFI). EPUBCFI attempts to solve the "right-hand-side" of a hyperlink citation: once the publication has been identified, where in the document is the citation found?

EPUB 3 is a new specification and no early implementations of EPUBCFI are available. A system for generating EPUBCFI citations out of an ebook (or for traversing them), perhaps combined with work out of the "left-hand-side" (publication identification) coming out of the NISO Ebook Annotations meetings [http://www.niso.org/topics/ccm/e-book_annotation/] and easily integrated with Zotero or other tools, could be tremendous valuable for scholars in elevating ebooks into first-class citizens in a research context.

3. Development of a pilot to produce ebooks with linked-data content

A cohort of Masters candidates could be engaged in a pilot program to produce ebooks with as much linked data [http://linkeddata.org/] content as possible. Linked data can be produced in an almost limitless number of schemas and representations. One of the goals of this pilot would be to find a "sweet spot"

in which linked data can be easily authored, and—critically—be used by end users in a useful way. Some cautionary notes should be taken from the recent memo by now-former TEI-C Board Chair Martin Mueller [http://ariadne.northwestern.edu/mmueller/teiletter.pdf], in which he warns:

> "…scholarly end users who encounter TEI-encoded texts almost never encounter them in an environment where they can take advantage of the distinct affordances of that encoding."

The value proposition of linked data content in this pilot should be clear to the authors; otherwise the process would simply be imposing an additional authorial burden.

4. Native EPUB output for Microsoft Word or Open Office

Apple Pages is the only major commercial word processor with native EPUB output. EPUB output from Pages [http://blog.threepress.org/2011/06/01/can-an-author-create-an-epub-file-using-normal-tools-part-1-pages/] can be high-quality, but it is tricky to author and not well-documented. Pages is also only available for the Macintosh and is not widely deployed on university computer systems nor is it easy to collaborate with Microsoft Word users.

A native Word export is badly needed in the ebook ecosystem, although the practical difficulties in creating clean markup from arbitrarily-structured Word documents should not be underestimated. This may be more feasible in the next-generation version of Microsoft Word, which will have a completely revamped HTML output.

Open Office, while not as widely-deployed as Microsoft Word, may be a good alternative. Plugins that support EPUB export do exist [http://extensions.services.openoffice.org/project/writer2xhtml], but native support is not a planned feature at this time [https://issues.apache.org/ooo/show_bug.cgi?id=104604]. A native export feature would be preferable to a plugin as it would be more rigorously tested across different platforms and less subject to becoming abandoned when the original maintainer moves on to other projects.

5. LaTeX to EPUB 3/MathML

Workflows that can accept LaTeX-based math but output valid EPUB 3 documents that include presentational MathML. This would allow mathematicians and scientists to continue to work in markup languages that are familiar to them but output documents that are compatible with the wider ebook

ecosystem.

Ideally this could build on the work of like pandoc [http://johnmacfarlane.net/pandoc/]. Pandoc already supports inline mathematical markup [http://johnmacfarlane.net/pandoc/README.html#math] with an EPUB output, but it does not convert to MathML or EPUB 3, nor does it fall back to using MathJax or even SVG math rendering for EPUB 2 compatibility.

6. Ereading systems with scholarly annotation systems

While designing the "perfect" academic hardware reader is a significant investment, a pilot ereader designed to run on a tablet computer could provide some of the affordances that students in the university Kindle DX pilot lacked. These affordances could in turn make digital publishing for researchers more attractive, as there would be a built-in audience receptive to their electronic monographs and papers.

Specifically, these features which are not suitable for electronic ink-based displays could be incorporated into a modern tablet computer such as the iPad:

1. Touch-based interaction (rather than the Kindle's joystick/button paradigm)
2. Methodologies to facilitate skimming and random access to documents rather than linear reading
3. Afforances for richer annotation, and highlighting.

Reports from students in the Kindle DX pilot studies suggest that such an ereading system might address their concerns with the hardware ereader. Kindle DX users in the Princeton pilot reported that color screens would be most useful for creating differently-colored highlights, rather than being useful for viewing color illustrations. The Reed College pilot report [http://web.reed.edu/cis/about/kindle_pilot/Reed_Kindle_report.pdf], specifically notes that:

> [The] iPhone uses a structure identical to that of the Kindle DX software but is much better suited to academic reading, because the iPhone's touch navigation and keyboard make location numbers and notes considerably easier to enter than on the Kindle DX.

This pilot pre-dated the release of iPad; the even faster iPad 2 offers the possibility of a powerful processor which could quickly render many pages and provide a valuable UI for skimming and drawing annotations. (The Kindle Fire tablet, released in the United States in November, 2011, does provide a color screen and

touch capability, but its 7-inch display is unsuitable for viewing PDFs. A program designed for academic use should utilize a 10-inch tablet such as an iPad.)

7. Community resources for individual scholars wishing to epublish

There are numerous blogs and support groups for fiction or general non-fiction authors who would like to pursue self- or digital-first publishing. There is no comparable resource for academic or scholarly publishers interested who may want to:

1. Understand the multiplicity of formats and approaches towards creating ebooks.
2. Experiment with form or with multimedia presentation of materials.
3. Develop methods of including interactive components, such as data visualizations or raw data components.
4. Learn how to format complex notation such as MathML or other domain-specific markup.
5. Discover where to submit manuscripts for digital publication, or how to host their own ebooks/e-publications.
6. Experiment with different ereading devices, smartphones and tablets in a physical lab.
7. Pool resources to perform QA and testing across different hardware and software platforms.

8. Organize training sessions on emerging ereading technologies such as EPUB 3.

Similarly, an organization to facilitate digital publications from small academic presses, similar to BiblioVault [http://www.bibliovault.org/] in the United States, but focused on UK presses and researchers.

Another organization to watch in this area is PressForward [http://pressforward.org/], organized under the banner of digital humanities in a web-centric way.

9. Maximize use of orphan works

Because the problem of orphan works (see Appendices) has special significance to academic researchers, addressing the identification of and fair use for such documents may be best done through the academic community. Indeed, in 2011 the University of Michigan cooperated with the HathiTrust to identify such works and make them available to the community if the rights allowed. Unfortunately

there was considerable resistance from the Authors Guild, who filed suit against the Trust [http://www.publishersweekly.com/pw/by-topic/digital/copyright/article/48659-authors-guild-sues-libraries-over-scan-plan.html]; subsequently the the orphan work list was found to be flawed [http://www.lib.umich.edu/news/u-m-library-statement-orphan-works-project]. There is opportunity to learn from the mistakes of the HathiTrust process and undertake a similar effort in the UK, provided that key stakeholders (including author representatives) can be part of the rights-holder identification and use process.

10. Community resources for institutions with digital collections

The time and resources required to undertake research on how researchers use digital media is considerable. To date there is no organization collecting this research and providing best-practices guidelines to institutions. This role may be particularly critical given the pace of change in digital publishing (for example, existing studies using the Kindle DX may already be obsolete as it is expected that the large-format e-ink device is soon to be discontinued [http://techcrunch.com/2011/11/27/kindle-dx-gets-temporary-price-cut-but-how-long-can-this-jumbo-e-reader-last/]). Another acute need is to understand the tradeoffs that users make between form factors: the convenience of mobile devices versus tablets with larger touchscreens versus full-featured laptops. Librarians, particularly in research libraries, require up-to-date recommendations to make to students and faculty, and need to be informed about usage patterns to make informed judgments about digital collection purchases.

APPENDICES

APPENDICES

Format and device reference tables

Ebook authoring tools

TABLE 6. EBOOK FORMAT COMPARISON

Format	Reflowable	Accessibility	Status	Fully internationalized?	Digital Rights Management
PDF	No	Low	Open	Yes	Many products
Mobipocket	Yes	Low	Proprietary	No	Amazon .AZW extension
EPUB 2	Yes	Medium	Open	Yes (1)	Yes, via Content Server (Adobe) and Fairplay (Apple)
EPUB 3	Yes	High	Open	Yes	None available yet

1.EPUB 2 supports the full Unicode set of characters but does not support scripts written vertically.

TABLE 7. EREADER DEVICE CAPABILITY COMPARISON

Device type	Screen type	Software stack	Color?	Video?	Battery life	Sunlight-readability	Text fidelity	Digital Rights Management
Dedicated EPUB ereader	Electronic ink	Adobe (1)	No	No	Excellent	Excellent	High	Various
Android tablet	LCD/ AMOLED	Android	Yes	Yes	Good	Fair	Moderate	Various
iPad	LCD	iOS	Yes	Yes	Very good	Fair	Moderate	Apple
iPhone 4	"Retina" LCD	iOS	Yes	Yes	Good	Fair	Moderate	Apple
Nook Color	LCD	Android/ Nook/ Adobe	Yes	Yes	Good	Fair	Moderate	Barnes & Noble
Kindle	Electronic ink	Linux/ Kindle	No	No	Excellent	Excellent	High	Amazon

1.Ereaders which license Adobe's Reader Mobile Software Development Kit (RMSDK) software for rendering EPUB books and decrypting DRM content include Nook, Kobo, Sony, Overdrive and others.

TABLE 8. EBOOK FORMAT DEVICE SUPPORT

Format	Kindle eink	Kindle apps	iBooks	Nook Color	Nook eink	Kobo apps
PDF	Kindle DX, others limited	None	Yes	Yes	Yes	Yes
Mobipocket (open)	All	All	No	No	Yes	No
Mobipocket (AZW)	All	All	No	No	No	No
EPUB 2	None (1)	None	Yes	Yes	Yes	Yes
EPUB 3 (2)	None	None	No	No	No	No

1.EPUB 2 files are accepted by the Kindle store for publishers and self-published authors. Users cannot add their own EPUB files directly.

2.EPUB 3 support is expected in devices and commercial ereading software by the first quarter of 2012.

Ebook authoring tools

Major desktop ebook creation software

These programs range from typical word processors with EPUB export to specialized ebook production systems.

Authoring system	Free/ Commercial	Audience	EPUB 2 output	PDF output	Mobi-pocket output	EPUB input	Notes
Apple Pages [http://www.apple.com/iwork/pages/]	Commercial	Authors, consumers, small businesses	Yes	Yes	No	No	Mac-only
Scrivener [http://www.literatureandlatte.com/scrivener.php]	Commercial	Authors, screenwriters	Yes	Yes	Yes	No	Mac-only (PC/Linux soon)
Adobe InDesign [http://www.adobe.com/products/indesign.html]	Commercial	Publishers	Yes	Yes	No	No	Page-focused, expensive
Atlantis [http://www.atlantiswordprocessor.com/en/]	Commercial	Authors	Yes	Yes	Yes	No	Windows-only
oXygen Author [http://www.oxygenxml.com/epub.html]	Commercial	Authors, publishers, XML editors	Yes	Yes	No	Yes	XML-centric
Calibre [http://calibre-ebook.com/]	Free	Ebook hobbyists	Yes	Yes	Yes	Yes	Conversion tool
Sigil [http://code.google.com/p/sigil/]	Free	Ebook hobbyists, micro-publishers	Yes	Yes	No	Yes	EPUB formatting tool

Note Excludes authoring systems which export PDF export only, or systems that rely on third-party plugins. There are no authoring systems for EPUB 3 at this time.

Blog or markup-based ebook creation systems

These solutions are released under a variety of licenses and may be hosted.

	License	Base platform	Source data	Output formats
Annotum [http://annotum.org/]	Open-source	WordPress	NLM/WYSIWYG	PDF, XML
Pressbooks [http://pressbooks.com/wp-signup.php]	Commercial, hosted	WordPress	HTML/WYSIWYG	EPUB, Mobi, PDF
Pandamian [http://www.pandamian.com/]	Commercial, hosted	Custom	HTML/WYSIWYG	EPUB, Mobi
Anthologize [http://anthologize.org/]	Open-source	WordPress	HTML/WYSIWYG	EPUB, PDF
asciidoc [http://www.methods.co.nz/asciidoc/index.html#_ebook_publication]	Open-source	Standalone	Simple text markup	EPUB, Mobi, PDF
pandoc [http://johnmacfarlane.net/pandoc/]	Open-source	Standalone	Various markup languages	PDF, EPUB, other markup formats
DocBook [http://www.docbook.org/]	Open-source	XSLT	DocBook XML	EPUB, Mobi

Note These are covered in depth in the jiscPUB blog as Making EPUB from WordPress (and other) web collections [http://jiscpub.blogs.edina.ac.uk/2011/05/25/making-epub-from-wordpress-and-other-web-collections/].

Glossary of ebook technical terms

CSS

Cascading Style Sheets, a method of adding visual styles such as bold or font choices to a digital document.

EPUB

The universal format for reflowable digital books.

EPUB 2

The version of the EPUB standard that is in wide industry use in 2011, suitable for most textual book content.

EPUB 3

Next-generation version of the EPUB specification, released in mid-2011, with enhancements to support multimedia and interactive content.

Kindle

An ecosystem of hardware, software, and digital books sold by Amazon, Inc.

Markup language

Text-based annotations in a document designed to inform computing systems of the underlying meaning or structure of text.

MathML

An XML format for expressing mathematical notation.

Mobipocket

An early reflowable ebook format, still the predominant format on the Amazon Kindle.

PDF

Portable Digital File, not actually portable to small device sizes and fixed to a particular page size.

Reflowable

Digital text that is not fixed to a particular screen size or ratio.

SVG

Scalable Vector Graphics, used to render line art in digital media.

XML

Text-based markup designed to inform computing systems of the underlying meaning or structure of text.